



# CAFÉ Distributed Computing Task Force

What is it we are going to talk about in this task force?

What do we need to achieve when?

Who will be working on it?

How do we organize ourselves?



# CAFÉ Distributed Computing Task Force

## Scope: Architectures for Distributed Computing

- ➔ Distributed Computing for Production: Simulation and Reconstruction
- ➔ Distributed Computing for Physics Analysis

## Distributed Computing Model: do we need to do it here?

- ➔ A) top down: Monarc, simulation etc; see Harvey's talk
- ➔ B) bottom up: how do we do things now, how could this evolve
  - ◆ Do we have good model for world-wide use of the (one) distributed CMS object database federation?
  - ◆ How will "users" use persistency of their objects, collections etc?
  - ◆ What environment will a user run "jobs" in?
  - ◆ How does a user find out which data exist/what this data is etc pp.
- ➔ Greg: some use cases and scenarios related to current production
- ➔ Some tough choices coming up
  - ◆ Everything (metadata) in the Objy database?
  - ◆ Job goes to data or data to job? etc pp.

## Iterative approach?!



# e.g. Production System

## Physicist wants to produce sample with special parameters

- ➔ Setup of the (remote?) environment
  - ◆ SCRAM, environment variables, cards files for generators, ORCA and calibration versions....
- ➔ Move, archive, catalog data files
  - ◆ File mover and catalogs, file archiving and storage caching, interfaces to ODMS

## Storing, cataloging and browsing, specifying “meta data”, querying CMS data

- ➔ Support physicists view on the data:
  - ◆ Which generator, cuts, cards, program versions, etc have been used
  - ◆ What was the cross section, luminosity

## Systems design issues

- ➔ Pile up simulation: does our current scheme scale? Can/do we need to we optimize it?
- ➔ Once the data is ready for physics:
  - ◆ Will we replicate everything everywhere? (No)
  - ◆ What is involved for a user to use a federation “somewhere else”?
  - ◆ Which data is where? Authentication? Remote environment?

## CAFÉ: understand, document and evaluate architectures



# Use Cases and Requirements

## Use Cases

- ➔ Are tasked to produce the relevant set of use cases
- ➔ “Tool” to measure function of system, quality, guide to develop architectures

## Scenarios

- ➔ Is it useful to start “just” looking at existing scenarios?
  - ◆ production systems (Fermilab, Italy, Caltech, ...)
  - ◆ Exporting user federation (Italy, Fermilab)
  - ◆ Tag queries (Caltech)
- ➔ This should tell as a lot about how the system works —> performance, bottlenecks
- ➔ Follow up on Dave’s use case:  
“A physicist at CERN uses Federation at Fermilab...”

## Requirements

- ➔ need to produce requirements document to the Grid Projects this summer
- ➔ Also: work patterns, virtual data, ...
- ➔ For this we need to understand the CMS distributed computing model, OR
  - ◆ A path how to approach a distributed computing model



# (incomplete) Action List for DCTF

## Use cases and computing model

- ➔ Based on use cases for analysis (Irwin) + production (to be taken care of here?)

## Document of what de-facto architectures exist

## evolve/iterate architectures towards “functional prototypes” (?)

- ➔ Production system
  - ◆ Including “using” the data in a “crude” way
- ➔ Query system
- ➔ Analysis system

## Grids: List what is needed (in terms of architecture definitions)

- ➔ File/replica and meta data catalogs
- ➔ Job submission, job environment, and job description
- ➔ Resource discovery and optimizations
- ➔ etc

## Mailing list!? Meeting date!?